



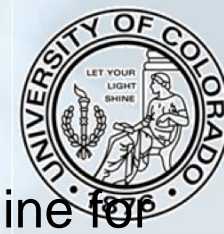
NCAR



Grid-BGC: a Grid-Enabled Research Platform for High-Resolution Surface Weather Interpolation and Biogeochemical Process

Nathan Wilhelmi, SCD/VETS

National Center for Atmospheric Research



Project Team

“Implementing an efficient supercomputer-based grid-compute engine for end-to-end operation of a high-resolution, high data-volume terrestrial carbon cycle model.”

Project PI: **Peter Thornton (NCAR)** Co-PI: **Henry Tufo (NCAR/CU)**

Staff:

Nathan Wilhelmi (NCAR)

Craig Hartsough (NCAR)

Matthew Woitaszek (CU)

Jason Cope (CU)

Collaborators:

Don Middleton (NCAR) Luca Cinquini (NCAR) Rich Loft (NCAR)



Project Funding

NASA has provided funding for the Grid-BGC project through the Advanced Information Systems Technology Office (NASA AIST Grant #NAG2-1646) and the Terrestrial Ecology Program.

University of Colorado computer time was provided by equipment purchased under DOE SciDAC Grant #DE-FG02-04ER63870, NSF ARI Grant #CDA-9601817, NSF sponsorship of the National Center for Atmospheric Research, and a grant from the IBM Shared University Research (SUR) program.



NCAR



Outline

- Project Overview
- Design Goals
- System Architecture
- User Interface
- Simulation Processing
- Visualization
- Grid Services / Job Execution
- Operational Testing
- Conclusions and Future work



NCAR



Research Problems

- Setting up and configuring simulations is complex, time consuming, and error prone.
- Researchers may not have the compute or storage resources available.
- The entire model workflow must be manually setup, executed, and managed.



NCAR



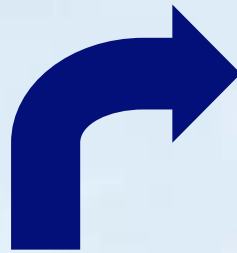
Purpose of GridBGC

- Use emerging Grid-Compute technologies to provide a research-quality platform for terrestrial carbon cycle modeling.
- Provide a Web Portal user interface to organize the complicated data dependencies that are typical of very large gridded ecosystem model implementations.
- Eliminate user interaction with remote computational resources by implementing automated job execution.
- Provide automated data streaming for model input and output datasets between the Portal, remote computational resources, and a remote mass storage facility.
- Provide robust analysis and visualization tools through the Portal.
- Demonstrate end-to-end functionality with a research-quality application (U.S. 1 km gridded simulations, targeting NACP).
- Focus on the needs of real researchers, through multiple iterations of platform development and beta-testing.

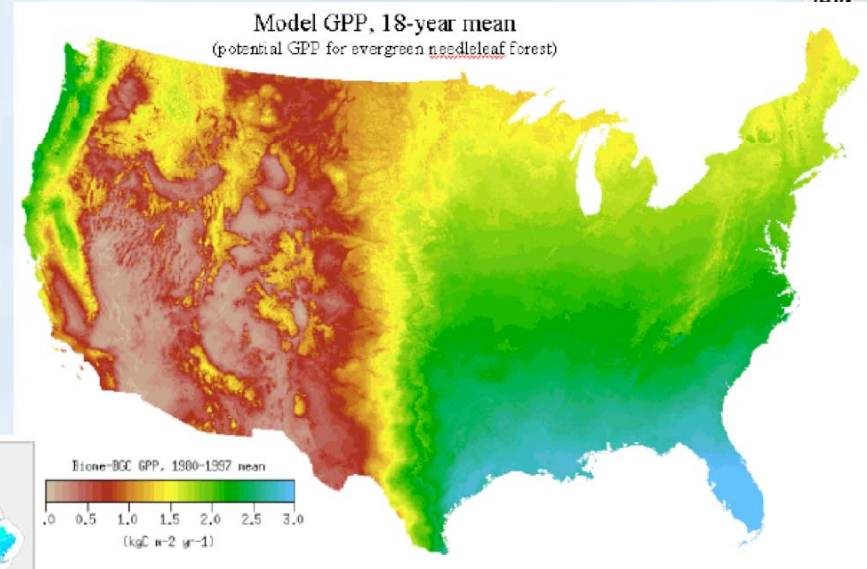
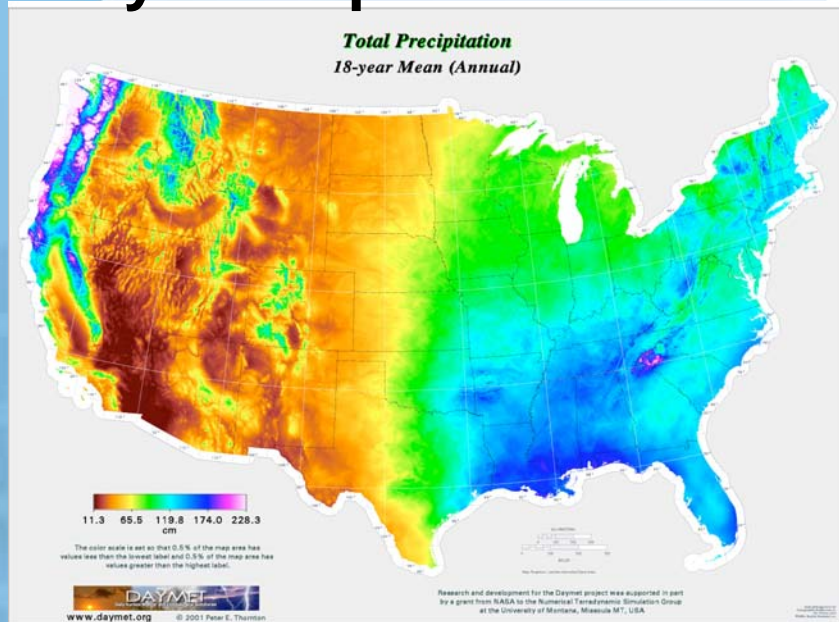
Science objective: large, gridded simulations of carbon cycle dynamics



Daymet – *interpolation and extrapolation of daily surface weather*



Daymet inputs...



...Grid-BGC outputs

BiomeBGC – *Carbon, nitrogen, and water cycles for land ecosystems*

Nathan Wilhelmi
wilhelmi@ucar.edu



NCAR



System Design Goals

- Focus on solving science problems.
- Abstract technology where ever possible.
- Support the typical modeling workflow for the Daymet and BiomeBGC models.
- Easy to configure and run simulations.
- Abstract data management where ever possible.



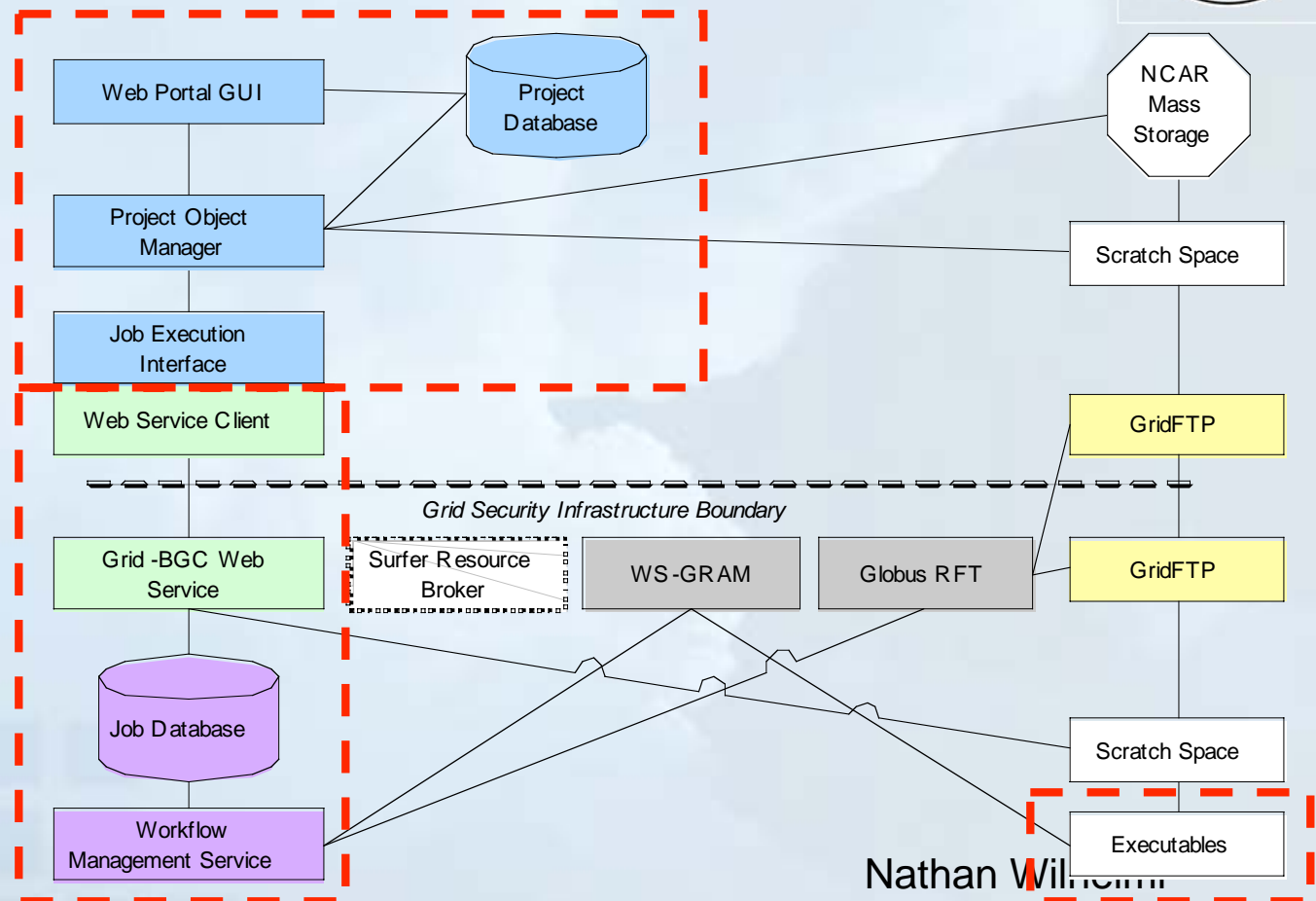
NCAR



System Architecture

Core Components:

- Graphical User Interface and support components
- Job execution services
- Science models



ESTC 2006
June 27 2006

Nathan Wilhelmi
wilhelmi@ucar.edu



NCAR



User Interface (Overview)

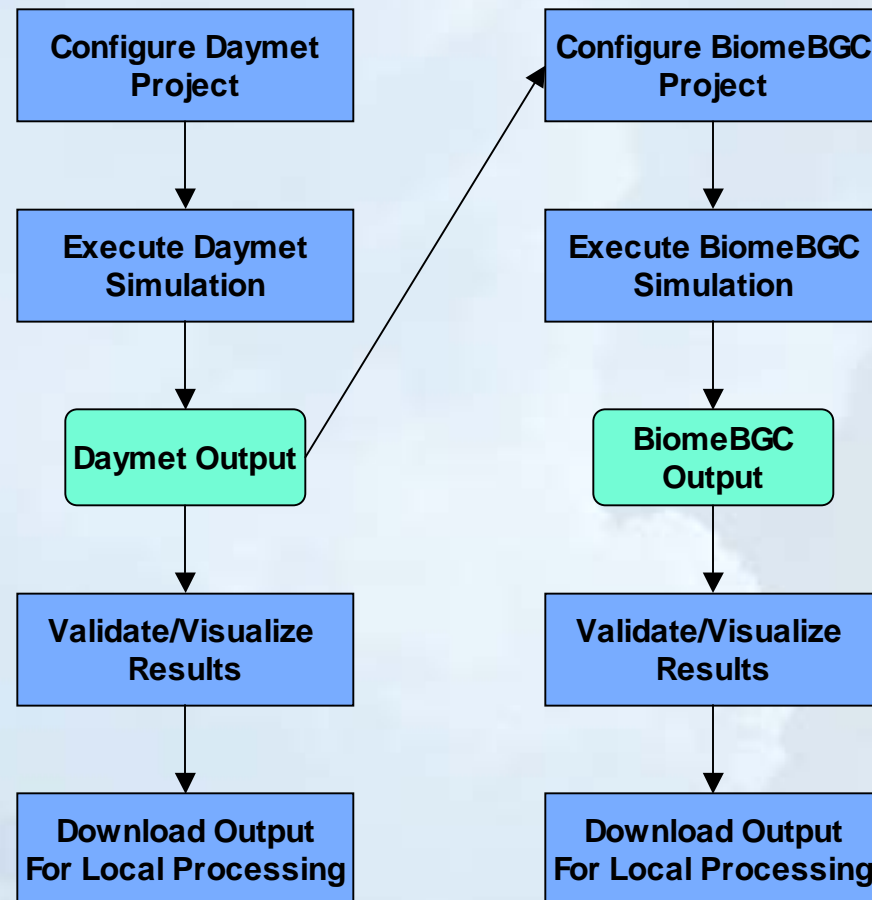
- Web based user interface.
- Developed and hosted at NCAR.
- Developed using common frameworks and tools:
 - Spring Framework
 - Hibernate
 - Globus COG Toolkit



NCAR



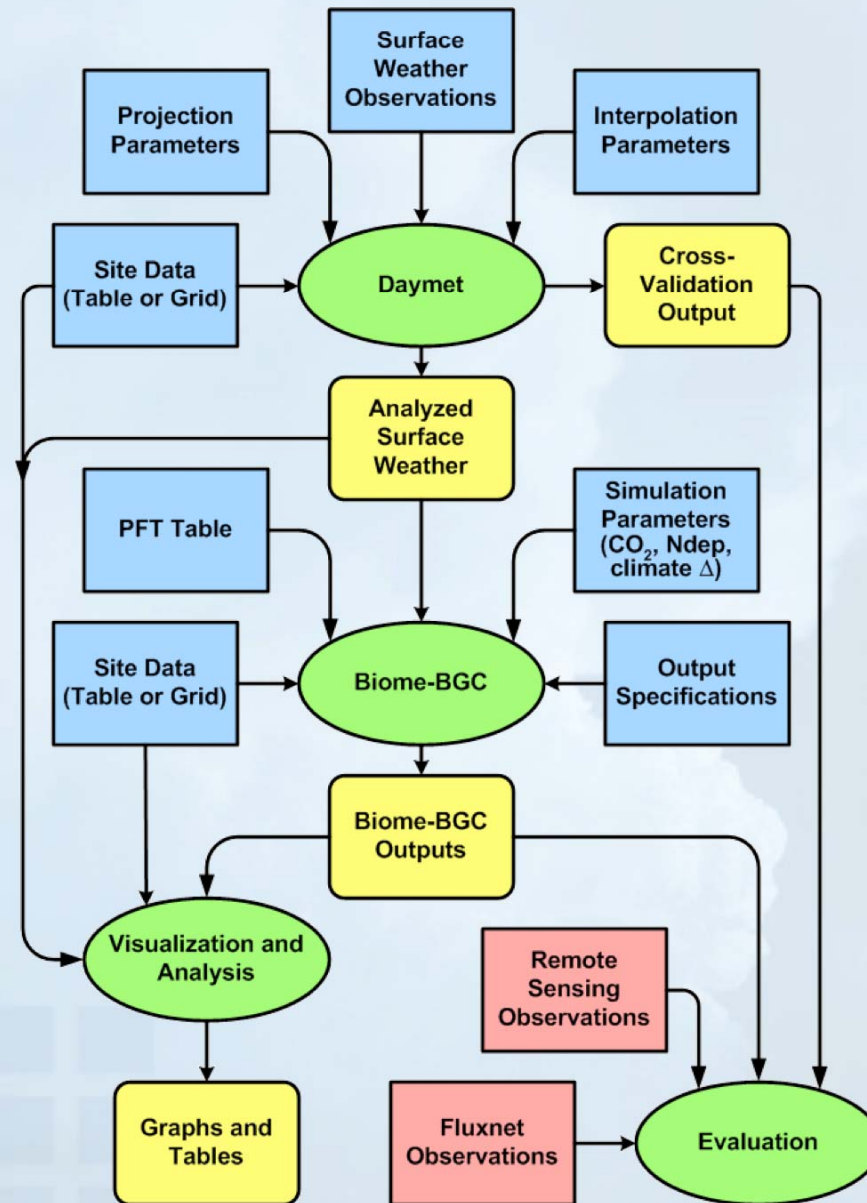
User Interface (Science Workflow)



User Interface (Project / Object Model)



NCAR



ESTC 2006
June 27 2006

Nathan Wilhelmi
wilhelmi@ucar.edu

User Interface (Site Data Object)

GridBGC Portal

WORKSPACE | LOG OFF | ABOUT | MODELS | CONTACT US

Current Daymet Project: Canada 1975 subset | Current BiomeBGC Spinup Project: 1x2 Spinup Project | Current BiomeBGC Project: Select... | Current User: wilhelmi

MY WORKSPACE **DAYMET** BIOME BGC VISUALIZATION ADMIN

Daymet Projects

- Daymet Projects
- Active
- Simulations
- Daymet
- Visualizations

Daymet Objects

- Surface
- Observations
- Projections
- Site Data (Grid)
- Simulation
- Parameters
- Simulation
- Output

Daymet Site Data Object Object Tasks

Object Security Settings...

This cannot be edited, it has been referenced.

This cannot be deleted, it has been referenced.

Assign to your current Daymet Project [Canada 1975 subset]...

Grid Site Data Object Details

Name: New Site Data Object

Description: Site Data Object Description

Grid Registration Values:

Number of Columns: 683

Number of Rows: 277

X - Lower Left Corner: -3345250.0

Y - Lower Left Corner: 165128.0

Cell Size (m): 10000.0

Tile Resolution (degrees): 20.0

Projection Details

Projection Not Specified.

Simulation Tile Details

TileId	North Edge	East Edge	South Edge	West Edge	Number of Pixels
113	50.0	-80.0	30.0	-100.0	1023
114	50.0	-60.0	30.0	-80.0	623
128	70.0	-140.0	50.0	-160.0	990
129	70.0	-120.0	50.0	-140.0	12453
130	70.0	-100.0	50.0	-120.0	14047
131	70.0	-80.0	50.0	-100.0	12054
132	70.0	-60.0	50.0	-80.0	16260
133	70.0	-40.0	50.0	-60.0	2452

Reference File Details

Description	Filename	File Size
Aspect Data	/datazone/gridBGC/storage/data/3519_aspect_data.nc	757280
Lat Degrees	/datazone/gridBGC/storage/data/3524_us_lat_degrees_data.nc	757252
Slope Data	/datazone/gridBGC/storage/data/3518_slope_data.nc	757280
East		

Done

dataportal.ucar.edu:2643

NCAR

UNIVERSITY OF COLORADO
LET YOUR LIGHT SHINE
1876

User Interface (Daymet Simulation Project)

Grid-BGC - Daymet Simulation Project Details - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

https://dataportal.ucar.edu:2643/GridBGCPortal/secure/daymetProjectView.htm?entityId=963

GridBGC Portal

WORKSPACE | LOG OFF | ABOUT | MODELS | CONTACT US

Current Daymet Project: Canada 1975 subset | Current BiomeBGC Spinup Project: 1x2 Spinup Project | Current BiomeBGC Project: Select... | Current User: wilhelmi

MY WORKSPACE **DAYMET** BIOME BGC VISUALIZATION ADMIN

Daymet Projects

- Daymet Projects
- Active
- Simulations
- Daymet
- Visualizations

Daymet Objects

- Surface
- Observations
- Projections
- Site Data (Grid)
- Simulation
- Parameters
- Simulation
- Output

Daymet Simulation Project Workflow Tasks

Step 1 : Simulation Configuration -> Status: LOCKED
 Step 2 : Simulation Run Control -> Terminate Simulation Wizard | Status: RUNNING [View Simulation Details]
 Step 3 : Visualization Post Processing

Daymet Simulation Project Workspace Status

Current Project: Set this project as your current Daymet Project...

Daymet Simulation Project Tasks

Daymet Simulation Project Details

Title:	canada 95 subset
Description:	corrected 1995, 50 stations withheld
Type:	
Surface Observation Dataset:	canada 95 subset
Daymet Site Data:	Canada 10km Good
Daymet Parameterization:	canada 95 subset

Derived Simulation Project Details

Number Of Years:	1 years
Model Resolution:	20.0

Projection Details

Name:	Canada 10 km
Description:	Canada 10 km grid projection
Projection Type:	Albers Conical Equal Area
Latitude of 1st Standard Parallel:	45.0
Latitude of 2nd Standard Parallel:	57.0
Longitude of Central Meridian:	-96.0
Latitude of Center of Projection:	0.0
False Easting:	0.0
False Northing:	0.0

Referenced Projection:

Copyright National Center For Atmospheric Research , 2004. All Rights Reserved.

Done

dataportal.ucar.edu:2643

NCAR



ni
lu



NCAR



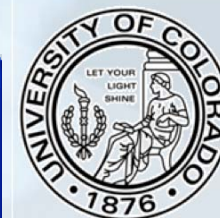
User Interface (Data Transformations)

- Models require specific NetCDF input files and model specific configurations.
 - Not produced by the typical tools used by researchers.
- Some datasets produced directly in the user interface.
- Users supply input datasets in easy to produce formats
 - ASCII Grid
 - ASCII Text Files
- Dataset are automatically converted into the required formats.
- Required derived datasets are automatically produced as required.

User Interface (Data Object Creation)



NCAR



GridBGC Portal

Current Daymet Project: Canada 1975 subset | Current BiomeBGC Spinup Project: 1x2 Spinup Project | Current BiomeBGC Project: Select... | Current User: wilhelmi

WORKSPACE | LOG OFF | ABOUT | MODELS | CONTACT US

MY WORKSPACE **DAYMET** BIOME BGC VISUALIZATION ADMIN

Daymet Projects

- Daymet Projects
- Active
- Simulations
- Daymet
- Visualizations

Daymet Objects

- Surface
- Observations
- Projections
- Site Data (Grid)
- Simulation
- Parameters
- Simulation
- Output

Create a New Grid Site Data Object

> Step 1 of 3 Enter object metadata.

Step 2 of 3 Enter dataset parameters and source dataset.

Step 3 of 3 Verify dataset and create new object.

Name: ESTC Presentation

Description: Testing

Creation Option:

☒ Create New Dataset

☐ Merge Existing

☐ Subset Existing

Next Cancel

Copyright National Center For Atmospheric Research , 2004. All Rights Reserved.

https://dataportal.ucar.edu:2643/GridBGCPortal/secure/Workspace.htm dataportal.ucar.edu:2643

ESTC 2006
June 27 2006

Nathan Wilhelmi
wilhelmi@ucar.edu

User Interface (Data Object Creation)



NCAR



GridBGC Portal

Current Daymet Project: Canada 1975 subset | Current BiomeBGC Spinup Project: 1x2 Spinup Project | Current BiomeBGC Project: Select... | Current User: wilhelmi

WORKSPACE | LOG OFF | ABOUT | MODELS | CONTACT US

MY WORKSPACE **DAYMET** BIOME BGC VISUALIZATION ADMIN

Daymet Projects

- Daymet Projects
- Active
- Simulations
- Daymet Visualizations

Daymet Objects

- Surface Observations
- Projections
- Site Data (Grid)
- Simulation Parameters
- Simulation Output

Create a New Grid Site Data Object

* Step 1 of 3 Enter object metadata.

> Step 2 of 3 Enter dataset parameters and source dataset.

Step 3 of 3 Verify dataset and create new object.

Name: ESTC Presentation

Description: Testing

Projection: Demo Projection

Archive Format: ZIP

Source Data Type: ESRI ASCII Grid

Site Data Archive to Upload: C:\projects\GridBGC\Da Browse...

Back Next Cancel

Copyright National Center For Atmospheric Research , 2004. All Rights Reserved.

Done dataportal.ucar.edu:2643

ESTC 2006
June 27 2006

Nathan Wilhelmi
wilhelmi@ucar.edu

User Interface (Data Object Creation)



GridBGC Portal

Current Daymet Project: Canada 1975 subset | Current BiomeBGC Spinup Project: 1x2 Spinup Project | Current BiomeBGC Project: Select... | Current User: wilhelmi

MY WORKSPACE DAYMET BIOME BGC VISUALIZATION ADMIN

Create a New Grid Site Data Object

- * Step 1 of 3 Enter object metadata.
- * Step 2 of 3 Enter dataset parameters and source dataset.
- > Step 3 of 3 Verify dataset and create new object.

Name: ESTC Presentation

Description: Testing

Projection: Display the projection parameters...

DEM Grid Values:

Number of Columns: 4587

Number of Rows: 2889

X - Lower Left Corner: -2050500.0

Y - Lower Left Corner: -2136500.0

Cell Size (m): 1000.0

Analysis Mask Grid Values:

Number of Columns: 4587

Number of Rows: 2889

X - Lower Left Corner: -2050500.0

Y - Lower Left Corner: -2136500.0

Cell Size (m): 1000.0

Back Finish Cancel

Copyright National Center For Atmospheric Research , 2004. All Rights Reserved.

ESTC 2006
June 27 2006

Nathan Wilhelmi
wilhelmi@ucar.edu

User Interface (Data Object Creation)



Grid-BGC - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

https://dataportal.ucar.edu:2643/GridBGCPortal/secure/SiteDataView.htm?entityId=965

GridBGC Portal

Current Daymet Project: Canada 1975 subset | Current BiomeBGC Spinup Project: 1x2 Spinup Project | Current BiomeBGC Project: Select... | Current User: wilhelmi

WORKSPACE | LOG OFF | ABOUT | MODELS | CONTACT US

MY WORKSPACE DAYMET BIOME BGC VISUALIZATION ADMIN

Daymet Projects

- Daymet Projects
- Active
- Simulations
- Daymet
- Visualizations

Daymet Objects

- Surface
- Observations
- Projections
- Site Data (Grid)
- Simulation
- Parameters
- Simulation
- Output

Daymet Site Data Object Object Tasks

Object Security Settings...
This cannot be edited, it has been referenced.
This cannot be deleted, it has been referenced.
Assign to your current Daymet Project [Canada 1975 subset]...

Grid Site Data Object Details

Name: ESTC Presentation
Description: Testing

Grid Registration Values:
Number of Columns: 4587
Number of Rows: 2889
X - Lower Left Corner: -2050500.0
Y - Lower Left Corner: -2136500.0
Cell Size (m): 1000.0
Tile Resolution (degrees): 2.0

Projection Details

Projection Not Specified.

Simulation Tile Details

TileId	North Edge	East Edge	South Edge	West Edge	Number of Pixels
11210	36.0	-80.0	34.0	-82.0	40512
11211	36.0	-78.0	34.0	-80.0	40504
11212	36.0	-76.0	34.0	-78.0	25935
11390	38.0	-80.0	36.0	-82.0	39504
11391	38.0	-78.0	36.0	-80.0	39504
11392	38.0	-76.0	36.0	-78.0	35155
11570	40.0	-80.0	38.0	-82.0	38433
11571	40.0	-78.0	38.0	-80.0	38430
11572	40.0	-76.0	38.0	-78.0	34898

Reference File Details

Description	Filename	File Size
Aspect Data	/datazone/gridBGC/storage/data/12983_aspect_data.nc	53007888
Lat Degrees	/datazone/gridBGC/storage/data/12988_us_lat_degrees_data.nc	53007860
Slope Data	/datazone/gridBGC/storage/data/12982_slope_data.nc	53007888
East Horizon	/datazone/gridBGC/storage/data/12985_east_horiz_data.nc	53007884

Done

dataportal.ucar.edu:2643

ESTC 2006
June 27 2006



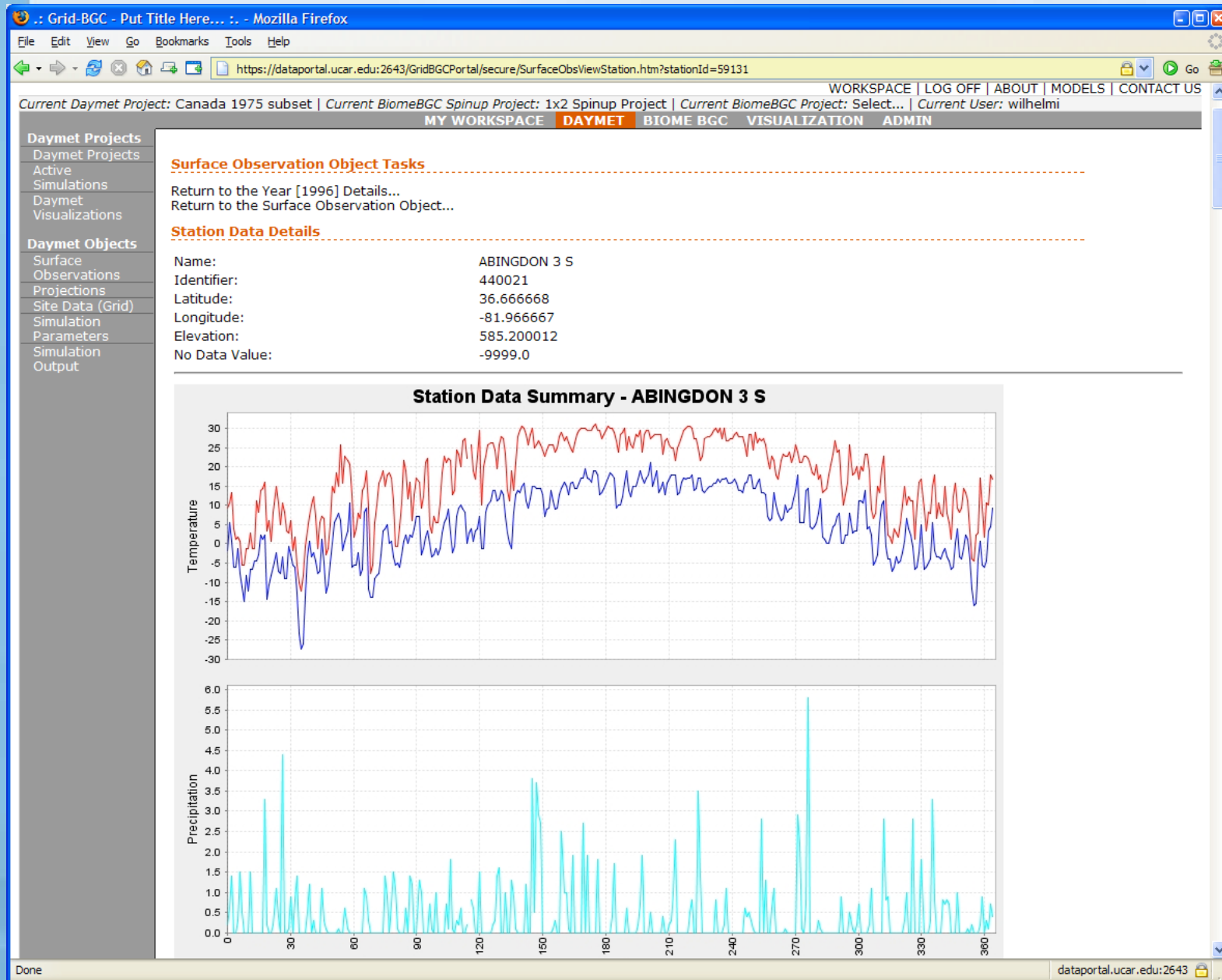
NCAR



User Interface (Quality Control)

- User supplied data validated where ever possible.
 - Required values
 - Data ranges
 - Missing data
- Project / Object model allows the user to only configure simulations that are compatible.
 - Data type
 - Spatial projection
- Visualizations used for data validation.

User Interface (Quality Control)



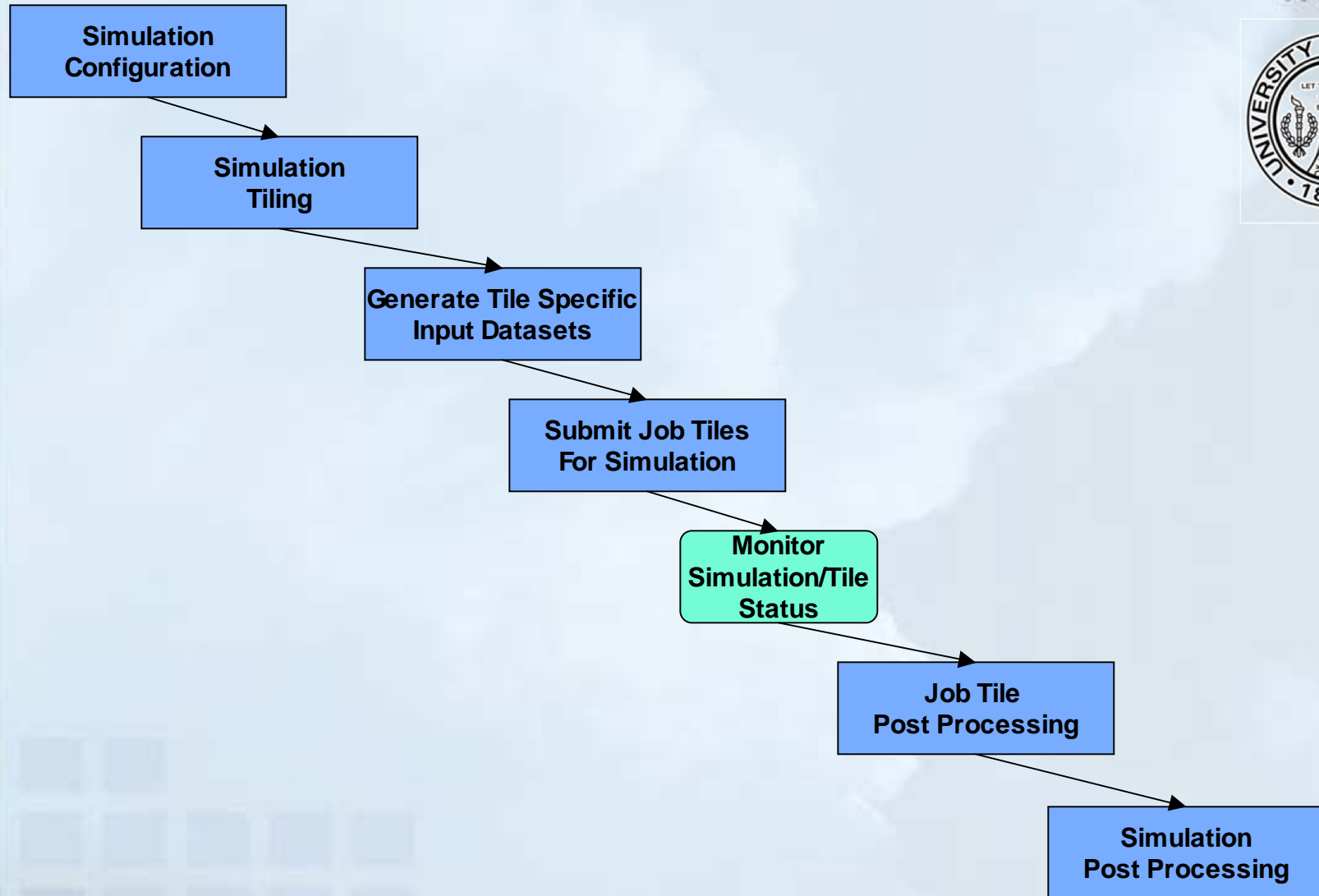
NCAR



Simulation Workflow



NCAR



ESTC 2006
June 27 2006

Nathan Wilhelmi
wilhelmi@ucar.edu



NCAR



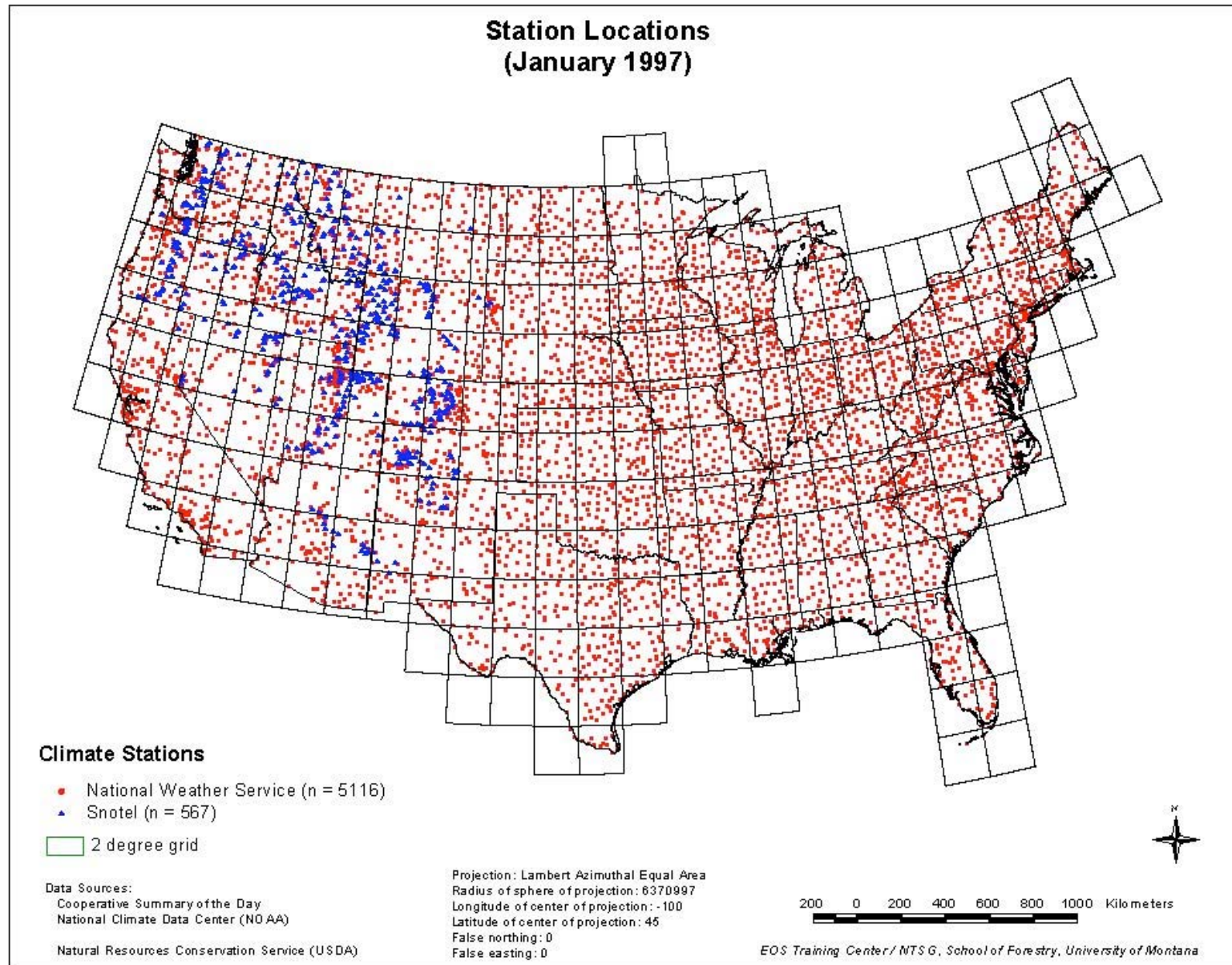
Job Tiling

- Simulations are divided into spatial tiles for computation.
- Job tiling process is driven by DEM resolution.
- Multiple simulation resolution supported.
- Daymet Simulations:
 - Tiles are computed for a single year for a spatial region.
- BiomeBGC Simulations:
 - Tiles are computed for all simulation years for a spatial region.

Job Tiling (Example)



CAR





NCAR



Job Monitoring and Control

- Job submission and status monitoring automated by the GUI.
- Status can be monitored on a both simulation and individual tile levels.

Job Monitoring and Control

GridBGC Portal

Current Daymet Project: Canada 1975 subset | Current BiomeBGC Spinup Project: 1x2 Spinup Project | Current BiomeBGC Project: Select... | Current User: wilhelmi

WORKSPACE | LOG OFF | ABOUT | MODELS | CONTACT US

MY WORKSPACE **DAYMET** BIOME BGC VISUALIZATION ADMIN

Daymet Projects

- Daymet Projects
- Active Simulations
- Daymet Visualizations

Daymet Objects

- Surface Observations
- Projections
- Site Data (Grid)
- Simulation Parameters
- Simulation Output

Daymet Simulation Run Status

Name: Canada 1975 subset
Status: STARTING -> ***RUNNING*** -> COMPLETE

TileId	Year	Pixels	Status	Statistics	JobTileId	RemoteJobId
111	1975	525	NOT_PROCESSED -> PRE_PROCESSING -> PENDING_SUBMISSION -> *SUBMITTED* -> STAGEIN -> EXECUTE -> STAGEOUT -> REMOTE_EXECUTION_COMPLETE -> POST_PROCESSING -> COMPLETE		1722	2899
112	1975	1616	NOT_PROCESSED -> PRE_PROCESSING -> PENDING_SUBMISSION -> *SUBMITTED* -> STAGEIN -> EXECUTE -> STAGEOUT -> REMOTE_EXECUTION_COMPLETE -> POST_PROCESSING -> COMPLETE		1723	2890
113	1975	4494	NOT_PROCESSED -> PRE_PROCESSING -> PENDING_SUBMISSION -> *SUBMITTED* -> STAGEIN -> EXECUTE -> STAGEOUT -> REMOTE_EXECUTION_COMPLETE -> POST_PROCESSING -> COMPLETE		1724	2891
114	1975	6804	NOT_PROCESSED -> PRE_PROCESSING -> PENDING_SUBMISSION -> *SUBMITTED* -> STAGEIN -> EXECUTE -> STAGEOUT -> REMOTE_EXECUTION_COMPLETE -> POST_PROCESSING -> COMPLETE		1725	2892
115	1975	998	NOT_PROCESSED -> PRE_PROCESSING -> PENDING_SUBMISSION -> *SUBMITTED* -> STAGEIN -> EXECUTE -> STAGEOUT -> REMOTE_EXECUTION_COMPLETE -> POST_PROCESSING -> COMPLETE		1726	2893
129	1975	7775	NOT_PROCESSED -> PRE_PROCESSING -> PENDING_SUBMISSION -> *SUBMITTED* -> STAGEIN -> EXECUTE -> STAGEOUT -> REMOTE_EXECUTION_COMPLETE -> POST_PROCESSING -> COMPLETE		1727	2894
130	1975	14330	NOT_PROCESSED -> PRE_PROCESSING -> PENDING_SUBMISSION -> *SUBMITTED* -> STAGEIN -> EXECUTE -> STAGEOUT -> REMOTE_EXECUTION_COMPLETE -> POST_PROCESSING -> COMPLETE		1728	2895
131	1975	10221	NOT_PROCESSED -> PRE_PROCESSING -> PENDING_SUBMISSION -> *SUBMITTED* -> STAGEIN -> EXECUTE -> STAGEOUT -> REMOTE_EXECUTION_COMPLETE -> POST_PROCESSING -> COMPLETE		1729	2896
132	1975	11991	NOT_PROCESSED -> PRE_PROCESSING -> PENDING_SUBMISSION -> *SUBMITTED* -> STAGEIN -> EXECUTE -> STAGEOUT -> REMOTE_EXECUTION_COMPLETE -> POST_PROCESSING -> COMPLETE		1730	2897
133	1975	1148	NOT_PROCESSED -> PRE_PROCESSING -> PENDING_SUBMISSION -> *SUBMITTED* -> STAGEIN -> EXECUTE -> STAGEOUT -> REMOTE_EXECUTION_COMPLETE -> POST_PROCESSING -> COMPLETE		1731	2898

Done

dataportal.ucar.edu:2643

NCAR





NCAR



Post-processing

- Individual tile post processing events
 - Daymet:
 - Extract cross validation statistics.
 - Generate summary time series visualizations.
 - BiomeBGC:
 - Generate summary time series visualizations.
- *Simulations post processing events*
 - *Mosaic process to combine tiled output into a contiguous dataset suitable for visualization.*



NCAR



Visualization

- *Under development.*
- Predefined visualization packages will be available.
- Visualization will be handled using the Unidata Integrated DataViewer (IDV).
<http://www.unidata.ucar.edu/software/idv/>
- Output data will be transformed into a compatible format.
- Direct access to visualization products will be through WebStart.



NCAR



Grid Services

- Computation services are exposed as explicit services for the both Daymet and BiomeBGC models.
- Globus Infrastructure Utilized
 - GT 4.0
 - MyProxy
 - GridFTP
 - WS-GRAM
 - RFT
 - Java COG Kit



NCAR



Job Execution Service

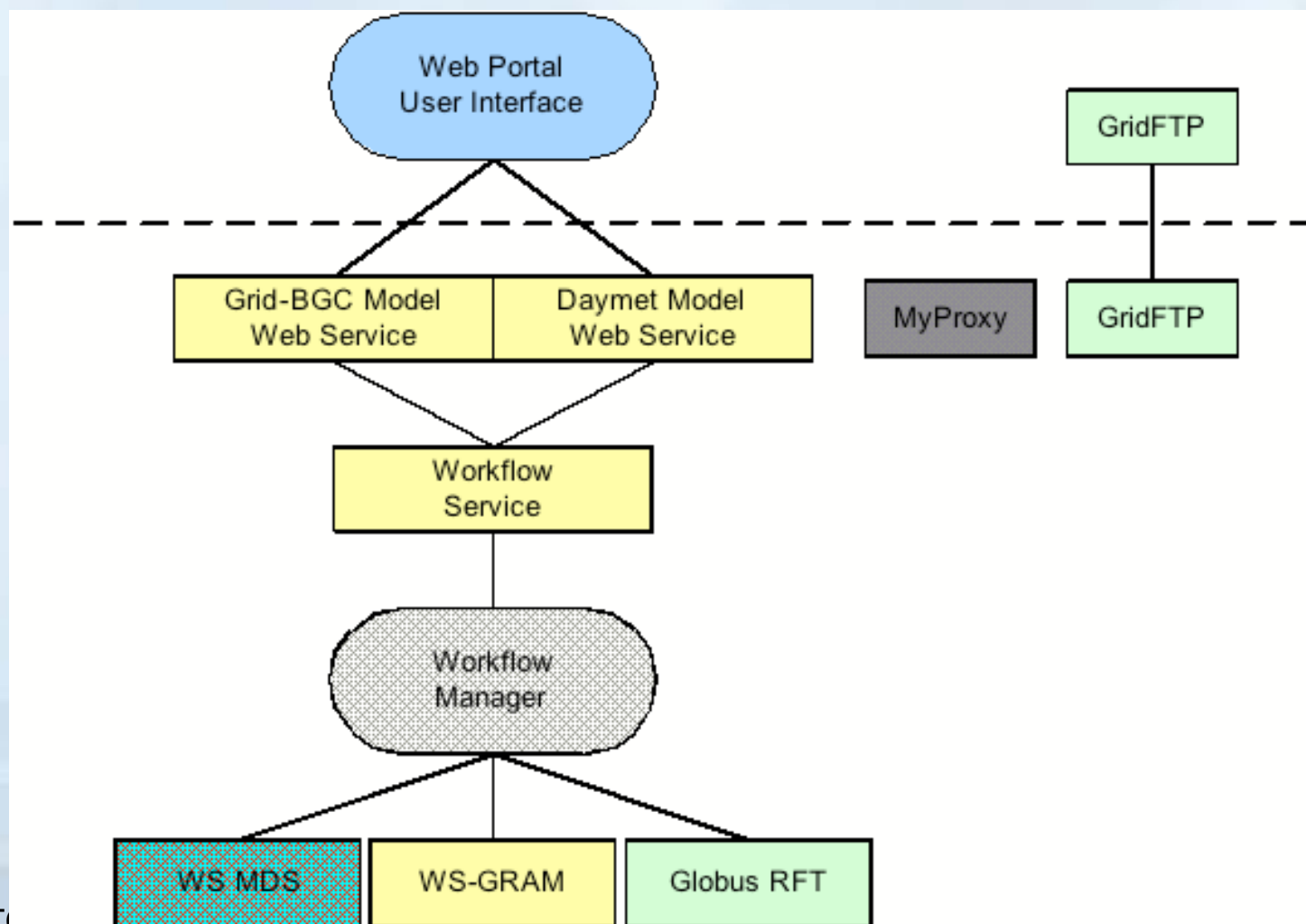
- Encapsulate legacy application setup, initialization, and submission operations inside web services
- Legacy application services become functional components and building blocks in a service-oriented architecture
- Reuse legacy application services in other grid-enabled applications or workflows



NCAR



Job Execution Service



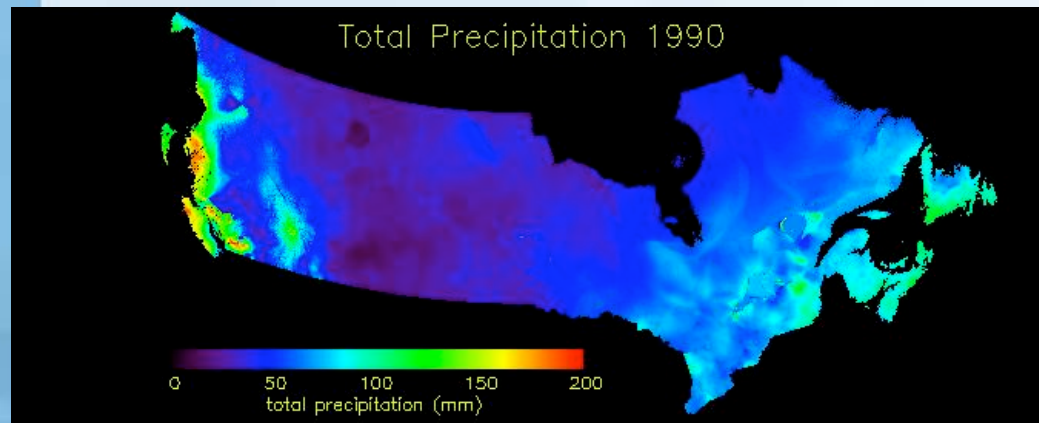
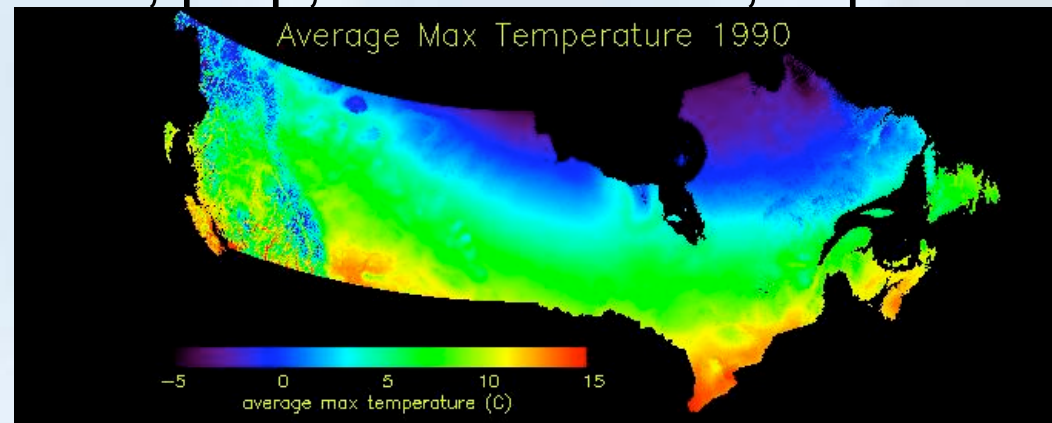


NCAR



Operational Testing – Canada

- 5 Variables: Tmax, Tmin, prcp, solar radiation, vapor pressure
- 10km Grid
- 1961-2003
- 6600 Stations
- 430 Jobs



June 27 2006

Nathan Wilhelmi
wilhelmi@ucar.edu



NCAR



Operational Testing

- Daymet
 - Complete U.S. run
 - 24 Years (1980-2003)
 - 1 km grid
- *BiomeBGC*
 - Canada, 10km grid.
 - Based on previously completed Daymet output
 - 2-3 years



NCAR



Future Work and Conclusions

- First operational tests were a success.
- Positive feedback received during testing.
- Job Execution Service has been demonstrated with other models.
- Automated data publishing to Community Data Portal
 - <http://cdp.ucar.edu>
- Establish additional computational resources
- Institutional policies appear to be the next barrier to address
 - Security policies
 - Data access
 - Resource Accounting